RAW SEQUENCE LISTING ERROR REPORT

The Biotechnology Systems Branch of the Scientific and Technical Information Center (STIC) detected errors when processing the following computer readable form:

Application Serial Number: 09/993,179

Source: OIPE

Date Processed by STIC: 12/5/2001

THE ATTACHED PRINTOUT EXPLAINS DETECTED ERRORS.
PLEASE FORWARD THIS INFORMATION TO THE APPLICANT BY EITHER:
1) **INCLUDING A COPY OF THIS PRINTOUT IN YOUR NEXT COMMUNICATION TO THE APPLICANT, WITH A NOTICE TO COMPLY or,**
2) **TELEPHONING APPLICANT AND FAXING A COPY OF THIS PRINTOUT, WITH A NOTICE TO COMPLY**

FOR CRF SUBMISSION QUESTIONS, PLEASE CONTACT MARK SPENCER, 703-308-4212.

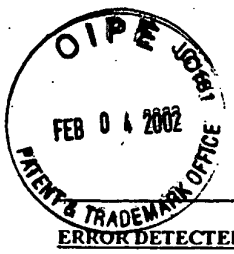
FOR SEQUENCE RULES INTERPRETATION, PLEASE CONTACT ROBERT WAX, 703-308-4216.
PATENTIN 2.1 e-mail help: patin21help@uspto.gov or phone 703-306-4119 (R. Wax)
PATENTIN 3.0 e-mail help: patin3help@uspto.gov or phone 703-306-4119 (R. Wax)

TO REDUCE ERRORED SEQUENCE LISTINGS, PLEASE USE THE CHECKER VERSION 3.0 PROGRAM, ACCESSIBLE THROUGH THE U.S. PATENT AND TRADEMARK OFFICE WEBSITE. SEE BELOW:

Checker Version 3.0

The Checker Version 3.0 application is a state-of the-art Windows based software program employing a logical and intuitive user-interface to check whether a sequence listing is in compliance with format and content rules. Checker Version 3.0 works for sequence listings generated for the original version of 37 CFR §§1.821 – 1.825 effective October 1, 1990 (old rules) and the revised version (new rules) effective July 1, 1998 as well as World Intellectual Property Organization (WIPO) Standard ST.25.
Checker Version 3.0 replaces the previous DOS-based version of Checker, and is Y2K-compliant. Checker allows public users to check sequence listings in Computer Readable form (CRF) before submitting them to the United States Patent and Trademark Office (USPTO). Use of Checker prior to filing the sequence listing is expected to result in fewer errored sequence listings, thus saving time and money.

Checker Version 3.0 can be down loaded from the USPTO website at the following address:
http://www.uspto.gov/web/offices/pac/checker



Raw Sequence Listing Error Summary

ERROR DETECTED	SUGGESTED CORRECTION	SERIAL NUMBER: 09/993,179

ATTN: NEW RULES CASES: PLEASE DISREGARD ENGLISH "ALPHA" HEADERS, WHICH WERE INSERTED BY PTO SOFTWAR

1	Wrapped Nucleics Wrapped Aminos	The number/text at the end of each line "wrapped" down to the next line. This may occur if your file was retrieved in a word processor after creating it. Please adjust your right margin to .3; this will prevent "wrapping."
2	Invalid Line Length	The rules require that a line not exceed 72 characters in length. This includes white spaces.
3 ✓	Misaligned Amino Numbering	The numbering under each 5th amino acid is misaligned. Do not use tab codes between numbers; use space characters, instead.
4	Non-ASCII	The submitted file was not saved in ASCII(DOS) text, as required by the Sequence Rules. Please ensure your subsequent submission is saved in ASCII text.
5	Variable Length	Sequence(s)____contain n's or Xaa's representing more than one residue. Per Sequence Rules, each n or Xaa can only represent a single residue. Please present the maximum number of each residue having variable length and indicate in the <220>-<223> section that some may be missing.
6	PatentIn 2.0 "bug"	A "bug" in PatentIn version 2.0 has caused the <220>-<223> section to be missing from amino acid sequences(s) _____. Normally, PatentIn would automatically generate this section from the previously coded nucleic acid sequence. Please manually copy the relevant <220>-<223> section to the subsequent amino acid sequence. This applies to the mandatory <220>-<223> sections for Artificial or Unknown sequences.
7	Skipped Sequences (OLD RULES)	Sequence(s) _____ missing. If intentional, please insert the following lines for each skipped sequence: (2) INFORMATION FOR SEQ ID NO:X: (insert SEQ ID NO where "X" is shown) (i) SEQUENCE CHARACTERISTICS: (Do not insert any subheadings under this heading) (xi) SEQUENCE DESCRIPTION:SEQ ID NO:X: (insert SEQ ID NO where "X" is shown) This sequence is intentionally skipped Please also adjust the "(ii) NUMBER OF SEQUENCES:" response to include the skipped sequences.
8	Skipped Sequences (NEW RULES)	Sequence(s) _____ missing. If intentional, please insert the following lines for each skipped sequence. <210> sequence id number <400> sequence id number 000
9	Use of n's or Xaa's (NEW RULES)	Use of n's and/or Xaa's have been detected in the Sequence Listing. Per 1.823 of Sequence Rules, use of <220>-<223> is MANDATORY if n's or Xaa's are present. In <220> to <223> section, please explain location of n or Xaa, and which residue n or Xaa represents.
10	Invalid <213> Response	Per 1.823 of Sequence Rules, the only valid <213> responses are: Unknown, Artificial Sequence, or scientific name (Genus/species). <220>-<223> section is required when <213> response is Unknown or is Artificial Sequence
11	Use of <220>	Sequence(s) _____ missing the <220> "Feature" and associated numeric identifiers and responses. Use of <220> to <223> is MANDATORY if <213> "Organism" response is "Artificial Sequence" or "Unknown." Please explain source of genetic material in <220> to <223> section. (See "Federal Register," 06/01/1998, Vol. 63, No. 104, pp. 29631-32) (Sec. 1.823 of Sequence Rules)
12	PatentIn 2.0 "bug"	Please do not use "Copy to Disk" function of PatentIn version 2.0. This causes a corrupted file, resulting in missing mandatory numeric identifiers and responses (as indicated on raw sequence listing). Instead, please use "File Manager" or any other manual means to copy file to floppy disk.
13	Misuse of n!	n can only be used to represent a single nucleotide in a nucleic acid sequence. N is not used to represent any value not specifically a nucleotide.

AMC/MH – Biotechnology Systems Branch – 08/21/2001

OIPE

RAW SEQUENCE LISTING DATE: 12/05/2001
PATENT APPLICATION: US/09/993,179 TIME: 09:54:30

Input Set : A:\sequence listing.txt
Output Set: N:\CRF3\11212001\I993179.raw

Does Not Comply
Corrected Diskette Needed

pp 1-3

3 <110> APPLICANT: McCarthy, Sean A.
4 Kuranda, Michael Joseph
5 Bulawa, Christine Ellen
6 Bossone, Steven
8 <120> TITLE OF INVENTION: METHOD FOR IDENTIFYING GENES ENCODING SIGNAL SEQUENCES
10 <130> FILE REFERENCE: 09404/032001
E--> 12 <140> CURRENT APPLICATION NUMBER: US/09/993,179
E--> 13 <141> CURRENT FILING DATE: 2001-11-06
15 <160> NUMBER OF SEQ ID NOS: 15
17 <170> SOFTWARE: FastSEQ for Windows Version 3.0

ERRORED SEQUENCES

52 <210> SEQ ID NO: 2
53 <211> LENGTH: 50
54 <212> TYPE: PRT
55 <213> ORGANISM: Homo sapiens
57 <400> SEQUENCE: 2
58 Met Lys Gly Thr Cys Val Ile Ala Trp Leu Phe Ser Ser Leu Gly Leu
E--> 59 1 5 10 15 10 15
60 Trp Arg Leu Ala His Pro Glu Ala Gln Gly Thr Thr Gln Cys Gln Arg
E--> 61 20 25 30
62 Thr Leu Glu Val Asn Ile Val Ser Pro Ser Ser Lys Ala Thr Phe Ser
E--> 63 35 40 45
64 Pro Ser
65 50

misabeled amino acid nos. (see Item 3 on Error Summary Sheet)

112 <210> SEQ ID NO: 4
113 <211> LENGTH: 125
114 <212> TYPE: PRT
115 <213> ORGANISM: Homo sapiens
117 <400> SEQUENCE: 4
118 Met Arg Ser Leu Leu Arg Thr Pro Phe Leu Cys Gly Leu Leu Trp Ala
E--> 119 1 5 10 15
120 Phe Cys Ala Pro Gly Ala Arg Ala Glu Glu Pro Ala Ala Ser Phe Ser
E--> 121 20 25 30
122 Gln Pro Gly Ser Met Gly Leu Asp Lys Asn Thr Val His Asp Gln Glu
E--> 123 35 40 45
124 His Ile Met Glu His Leu Glu Gly Val Ile Asn Lys Pro Glu Ala Glu
E--> 125 50 55 60
126 Met Ser Pro Gln Glu Leu Gln Leu His Tyr Phe Lys Met His Asp Tyr
E--> 127 65 70 75 80
128 Asp Gly Asn Asn Leu Leu Asp Gly Leu Glu Leu Ser Thr Ala Ile Thr
E--> 129 85 90 95
130 His Val His Lys Glu Glu Gly Ser Glu Gln Ala Pro Leu Glu Val Asn
E--> 131 100 105 110
132 Ile Val Ser Pro Ser Ser Lys Ala Thr Phe Ser Pro Ser

same error

RAW SEQUENCE LISTING DATE: 12/05/2001
PATENT APPLICATION: US/09/993,179 TIME: 09:54:30

Input Set : A:\sequence listing.txt
Output Set: N:\CRF3\11212001\I993179.raw

```
E--> 133    115       120   125
     135 <210> SEQ ID NO: 5
     136 <211> LENGTH: 32
     137 <212> TYPE: PRT
     138 <213> ORGANISM: Mus musculus
     140 <400> SEQUENCE: 5
     141 Met Lys Gly Ala Cys Ile Leu Ala Trp Leu Phe Ser Ser Leu Gly Val
E--> 142   1     5         10    15
     143 Trp Arg Leu Ala Arg Pro Glu Thr Gln Asp Pro Ala Lys Cys Gln Arg
E--> 144         20    25        30
     146 <210> SEQ ID NO: 6
     147 <211> LENGTH: 45
     148 <212> TYPE: PRT
     149 <213> ORGANISM: Homo sapiens
     151 <400> SEQUENCE: 6
     152 Met Ser Pro Gln Glu Leu Gln Leu His Tyr Phe Lys Met His Asp Tyr
E--> 153   1     5         10    15
     154 Asp Gly Asn Asn Leu Leu Asp Gly Leu Glu Leu Ser Thr Ala Ile Thr
E--> 155         20    25        30
     156 His Val His Lys Glu Glu Gly Ser Glu Gln Ala Pro Leu
E--> 157       35        40    45
     238 <210> SEQ ID NO: 14
     239 <211> LENGTH: 32
     240 <212> TYPE: PRT
     241 <213> ORGANISM: Homo sapiens
     243 <400> SEQUENCE: 14
     244 Met Lys Gly Thr Cys Val Ile Ala Trp Leu Phe Ser Ser Leu Gly Leu
E--> 245   1     5         10    15
     246 Trp Arg Leu Ala His Pro Glu Ala Gln Gly Thr Thr Gln Cys Gln Arg
E--> 247         20    25        30
     249 <210> SEQ ID NO: 15
     250 <211> LENGTH: 108
     251 <212> TYPE: PRT
     252 <213> ORGANISM: Homo sapiens
     254 <400> SEQUENCE: 15
     255 Met Arg Ser Leu Leu Arg Thr Pro Phe Leu Cys Gly Leu Leu Trp Ala
E--> 256   1     5         10    15
     257 Phe Cys Ala Pro Gly Ala Arg Ala Glu Glu Pro Ala Ala Ser Phe Ser
E--> 258         20    25        30
     259 Gln Pro Gly Ser Met Gly Leu Asp Lys Asn Thr Val His Asp Gln Glu
E--> 260       35        40    45
     261 His Ile Met Glu His Leu Glu Gly Val Ile Asn Lys Glu Ala Glu Met
E--> 262       50    55        60
     263 Ser Pro Gln Glu Leu Gln Leu His Tyr Phe Lys Met His Asp Tyr Asp
E--> 264  65         70    75        80
     265 Gly Asn Asn Leu Leu Asp Gly Leu Glu Leu Ser Thr Ala Ile Thr His
E--> 266          85        90    95
     267 Val His Lys Glu Glu Gly Ser Glu Gln Ala Pro Leu
E--> 268        100   105
```

same

same

same

same

(see next page)

<400> 1

```
ggggaccgtg tttgtggccc ccaagccggt gccccccatt ttggaactca gcgagtaggg        60
ggcggctctg gggaagtggc agggggcgca gcagctgctg cctccacttc cctagccagg       120
tgctgaagag gatcttcgga gccgctctgg cccccaggcg ctggatgact ggcaccagcg       180
ctcctcgcac ctgtgttggt gtgtgagact tgggctggag tgcccacgtg gctgtggagt       240
cagtgtgatt catgattgag gaaacgcgtc ctccatcctc tctctccttg gcactttcca       300
cacatgagga gaagaagagc ttctgtttag aagacacgtg cccagagtca gaggccccct       360
gcccacc atg aag gga acc tgt gtt ata gca tgg ctg ttc tca agc ctg   409
      Met Lys Gly Thr Cys Val Ile Ala Trp Leu Phe Ser Ser Leu
        1               5                  10

ggg ctg tgg aga ctc gcc cac cca gag gcc cag ggt acg act cag tgc   457
Gly Leu Trp Arg Leu Ala His Pro Glu Ala Gln Gly Thr Thr Gln Cys
 15                  20                  25                  30

cag aga aca ctc gag gtg aat att gtt tcc ccc agc tcc aag gca aca   505
Gln Arg Thr Leu Glu Val Asn Ile Val Ser Pro Ser Ser Lys Ala Thr
                 35                  40                  45

ttc agt cca agt                                                  517
Phe Ser Pro Ser
           50
```

move amino acid directly under respective amino acid

(please correct this misalignment in subsequent coding sequence)

FYI Use of n and/or Xaa has been detected in the Sequence Listing. Review the Sequence Listing to insure a corresponding explanation is presented in the <220> to <223> fields of each sequence using n or Xaa.

VERIFICATION SUMMARY DATE: 12/05/2001
PATENT APPLICATION: US/09/993,179 TIME: 09:54:31

Input Set : A:\sequence listing.txt
Output Set: N:\CRF3\11212001\I993179.raw

L:12 M:270 C: Current Application Number differs, Replaced Current Application Number
L:13 M:271 C: Current Filing Date differs, Replaced Current Filing Date
L:41 M:336 W: Invalid Amino Acid Number in Coding Region, SEQ ID:1
L:45 M:336 W: Invalid Amino Acid Number in Coding Region, SEQ ID:1
L:49 M:336 W: Invalid Amino Acid Number in Coding Region, SEQ ID:1
L:59 M:332 E: (32) Invalid/Missing Amino Acid Numbering, SEQ ID:2
M:332 Repeated in SeqNo=2
L:81 M:336 W: Invalid Amino Acid Number in Coding Region, SEQ ID:3
L:85 M:336 W: Invalid Amino Acid Number in Coding Region, SEQ ID:3
L:89 M:336 W: Invalid Amino Acid Number in Coding Region, SEQ ID:3
L:93 M:336 W: Invalid Amino Acid Number in Coding Region, SEQ ID:3
L:97 M:336 W: Invalid Amino Acid Number in Coding Region, SEQ ID:3
L:101 M:336 W: Invalid Amino Acid Number in Coding Region, SEQ ID:3
L:105 M:336 W: Invalid Amino Acid Number in Coding Region, SEQ ID:3
L:109 M:336 W: Invalid Amino Acid Number in Coding Region, SEQ ID:3
L:119 M:332 E: (32) Invalid/Missing Amino Acid Numbering, SEQ ID:4
M:332 Repeated in SeqNo=4
L:142 M:332 E: (32) Invalid/Missing Amino Acid Numbering, SEQ ID:5
M:332 Repeated in SeqNo=5
L:153 M:332 E: (32) Invalid/Missing Amino Acid Numbering, SEQ ID:6
M:332 Repeated in SeqNo=6
L:209 M:257 W: Feature value mis-spelled or invalid, <221> Name/Key for SEQ ID#:11
L:214 M:341 W: (46) "n" or "Xaa" used, for SEQ ID#:11
L:245 M:332 E: (32) Invalid/Missing Amino Acid Numbering, SEQ ID:14
M:332 Repeated in SeqNo=14
L:256 M:332 E: (32) Invalid/Missing Amino Acid Numbering, SEQ ID:15
M:332 Repeated in SeqNo=15